Comment on 'Analysis of Longitudinal Trials with Protocol Deviations — a Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation' by Carpenter, Roger and Kenward.

Carpenter et al. (2013) propose a multiple imputation (MI) approach for analysing data from clinical trials with protocol deviations. Sensitivity analysis to departures from MAR is widely acknowledged as important, but is poorly handled in practice, so we welcome their detailed proposals. However, here we highlight two problems with their method: an implicit assumption of non-informative deviation and failure of the Rubin's Rule (RR) variance estimator.

The method of Carpenter et al. (2013)

We start by summarising the method of Carpenter et al. (2013), using their notation and additional notation $\boldsymbol{\mu}_T$, $\boldsymbol{\mu}_{T,O}$, $\boldsymbol{\mu}_{T,M}$, $\boldsymbol{\Sigma}_{T,OO}$, $\boldsymbol{\Sigma}_{T,MO}$, $\boldsymbol{Y}_M^*$ and $\boldsymbol{Y}^*$. The number of repeated outcomes per patient and number of patients are J and n, respectively. For each patient, D denotes the deviation time (i.e. time of last outcome before protocol deviation), T is the randomisation group (r for reference, a for active), and $\boldsymbol{Y}_O$ are the outcomes prior to deviation. Let $\boldsymbol{Y}^* = (\boldsymbol{Y}_O^T, \boldsymbol{Y}_M^{*T})^T$, where $\boldsymbol{Y}_M^*$ denotes a vector of hypothetical outcomes after deviation. These may or may not be the same as the actual post-deviation outcomes $\boldsymbol{Y}_M$. Carpenter et al. specify separate normal distributions for $\boldsymbol{Y}^*$ given $T = r$ and for $\boldsymbol{Y}^*$ given $T = a$, and denote the unknown means of these distributions by $\boldsymbol{\mu}_r = (\mu_{r,1}, \ldots, \mu_{r,J})$ and $\boldsymbol{\mu}_a = (\mu_{a,1}, \ldots, \mu_{a,J})$, and the variances by $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_a$. Let $\boldsymbol{\mu}_{T,O}$ and $\boldsymbol{\mu}_{T,M}$ $(T = r, a)$ denote $(\mu_{T,1}, \ldots, \mu_{T,D})^T$ and $(\mu_{T,D+1}, \ldots, \mu_{T,J})^T$, respectively, and let the submatrices of $\boldsymbol{\Sigma}_T$ corresponding to $\mathrm{Var}(\boldsymbol{Y}_O \mid T)$ and $\mathrm{Cov}(\boldsymbol{Y}_M^*, \boldsymbol{Y}_O \mid T)$ be denoted $\boldsymbol{\Sigma}_{T,OO}$ and $\boldsymbol{\Sigma}_{T,MO}$, respectively. Carpenter et al. denoted $\boldsymbol{\Sigma}_{r,OO}$, $\boldsymbol{\Sigma}_{r,MO}$, $\boldsymbol{\Sigma}_{a,OO}$ and $\boldsymbol{\Sigma}_{a,MO}$ as, respectively, $\boldsymbol{R}_{11}$, $\boldsymbol{R}_{21}$, $\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{21}$. A non-informative prior is assumed for $(\boldsymbol{\mu}_r, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_a)$ and its posterior obtained under the assumption that the missingness mechanism is ignorable.

Under the assumption of 'randomised-arm MAR', the posterior predictive distribution

of the actual post-deviation outcomes $\boldsymbol{Y}_M$ is the same as that of $\boldsymbol{Y}_M^*$, and so can be multiply imputed using this distribution. Therefore, as described by Carpenter et al., imputation under 'randomised-arm MAR' is done by sampling a value of $(\boldsymbol{\mu}_r, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_a)$ from its posterior and then sampling $\boldsymbol{Y}_M$ from a normal distribution with mean $\boldsymbol{\mu}_{T,M} + \boldsymbol{\Sigma}_{T,MO}\boldsymbol{\Sigma}_{T,OO}^{-1}(\boldsymbol{Y}_O - \boldsymbol{\mu}_{T,O})$ and variance given by Carpenter et al. As an addition to this established MI procedure for randomised-arm MAR, Carpenter et al. propose four novel MI procedures for MNAR data. These procedures differ from that described for randomised-arm MAR in the mean and variance of the normal distribution from which $\boldsymbol{Y}_M$ is sampled. For 'jump to reference', the mean is $\boldsymbol{\mu}_{r,M} + \boldsymbol{\Sigma}_{r,MO}\boldsymbol{\Sigma}_{T,OO}^{-1}(\boldsymbol{Y}_O - \boldsymbol{\mu}_{r,O})$, for 'copy reference' it is $\boldsymbol{\mu}_{r,M} + \boldsymbol{\Sigma}_{r,MO}\boldsymbol{\Sigma}_{r,OO}^{-1}(\boldsymbol{Y}_O - \boldsymbol{\mu}_{r,O})$, for 'copy increments in reference' it is $(\mu_{T,D} + \mu_{r,D+1} - \mu_{r,D}, \ldots, \mu_{T,J} + \mu_{r,J} - \mu_{r,D})^T + \boldsymbol{\Sigma}_{r,MO}\boldsymbol{\Sigma}_{T,OO}^{-1}(\boldsymbol{Y}_O - \boldsymbol{\mu}_{T,O})$, and for 'last mean carried forward' (LMCF) it is $(\mu_{T,D}, \ldots, \mu_{T,D})^T + \boldsymbol{\Sigma}_{T,MO}\boldsymbol{\Sigma}_{T,OO}^{-1}(\boldsymbol{Y}_O - \boldsymbol{\mu}_{T,O})$.

Let $\hat{\theta}_q$ denote the treatment effect estimate from the qth imputed dataset ($q = 1, \ldots, Q$), and $\hat{\mathrm{Var}}(\hat{\theta}_q)$ be its variance estimate. The Q effect estimates are combined into an overall estimate $\hat{\theta}_{(Q)}$ using RR for the mean: $\hat{\theta}_{(Q)} = Q^{-1}\sum_{q=1}^{Q}\theta_q$. RR for the variance gives an estimate of the repeated sampling variance of $\hat{\theta}_{(Q)}$: $\hat{\mathrm{Var}}(\hat{\theta}_{(Q)}) = B_Q + (1 + Q^{-1})W_Q$, where $B_Q = Q^{-1}\sum_{q=1}^{Q}\hat{\mathrm{Var}}(\hat{\theta}_q)$ and $W_Q = (Q-1)^{-1}\sum_{q=1}^{Q}(\hat{\theta}_q - \hat{\theta}_{(Q)})^2$.

Problem 1: Informative deviations

The first problem with the procedures proposed by Carpenter et al. is that they make an implicit 'non-informative deviation' assumption, $P(D = t \mid D \geq t, T, \boldsymbol{Y}) = P(D = t \mid D \geq t, T, Y_1, \ldots, Y_D)$, i.e. that the hazard of deviation does not depend on later outcomes given earlier outcomes. For simplicity of exposition, suppose $J = 2$, there are no deviations in the reference group, and outcomes at different times are independent and the imputer knows this (however, the problem we now describe applies more generally). Under the 'jump to reference' and 'copy reference' assumptions, the mean of the imputation distribution of post-deviation Y_2 given deviation is $\mu_{r,2}$, which is the unconditional expected outcome in a randomly sampled untreated patient. This is a reasonable assumption if the factors influencing deviation are independent of those in-

fluencing Y_2. However, this will often not be the case. The following example illustrates what happens when deviation is informative.

For each patient, let D^* denote the (possibly counterfactual) time that the patient would have deviated had she/he been randomised to the active group. So, $D^* = D$ if $T = a$ and is missing if $T = r$. Suppose that $E(Y_2 \mid D^*, T) = \alpha + \beta I(D^* = 1)$. So, treatment has no effect on outcome, but outcomes of patients who deviate are, on average, greater by β than those who do not. Assume deviation is informative, i.e. $\beta \neq 0$. Let $\pi = P(D^* = 1) > 0$. The expected mean of the imputation distribution for post-deviation outcomes is $\mu_{r,2} = E(Y_2 \mid T) = \alpha + \beta\pi$, which is different from the true mean $E(Y_2 \mid D^* = 1, T) = \alpha + \beta$. Therefore, in the imputed dataset the mean of Y_2 in the active group has expectation $\pi(\alpha + \beta\pi) + (1 - \pi)\alpha = \alpha + \beta\pi^2$. This is different from $\alpha + \beta\pi$, the expected mean in the reference group, and so the treatment effect estimate is biased away from zero. Similar considerations apply in the case of 'copy increments in reference' and LMCF.

Problem 2: Use of the Rubin's Rule variance estimator

The second problem is that the RR estimator of the repeated sampling variance of $\hat{\theta}_{(Q)}$ may not be valid unless the data are 'randomised-arm MAR' and MI is carried out assuming this. This is because under the other missingness assumptions ('jump to reference', etc.), the imputer assumes more than the analyst, which is known to cause the RR variance estimator to overestimate the repeated sampling variance (Meng, 1994). The following extreme example illustrates this.

Assume non-informative deviation (so Problem 1 does not apply), $J = 2$, no deviation in the reference group, all patients in the active arm deviate at time 1 ($D = 1$), and outcomes at different times are independent and the imputer knows this. Suppose the treatment effect of interest is $\theta = E(Y_2 \mid T = a) - E(Y_2 \mid T = r)$ and the complete-data estimator of this effect is just the difference between the sample means in the two arms. The posterior of $\mu_{r,2}$ is normal with mean equal to the sample mean of Y_2 in the reference arm. Therefore, under 'jump to reference' or 'copy reference', $\hat{\theta}_q$

is normally distributed with mean zero. Consequently, $\hat{\theta}_{(\infty)} = 0$ and the repeated sampling variance of $\hat{\theta}_{(\infty)}$ equals zero. On the other hand, B_∞ and hence $\hat{\text{Var}}(\hat{\theta}_{(Q)})$ are both positive. The variance estimator is overestimating the true variance because the data are imputed under a strong assumption that is no longer made when these imputed data are analysed, specifically that there is no treatment effect in those who deviate.

More generally in the four MNAR imputation procedures, the imputer (but not the analyst) assumes a relation between the expected post-deviation outcomes of an individual in the active arm given that he/she deviates and the expected outcomes of an individual in the reference arm. This enables the imputer to use data from the reference arm when imputing post-deviation outcomes in the active arm. In 'randomised-arm MAR' imputation, on the other hand, the imputer does not assume a relation between outcomes in the two arms, and imputes post-deviation outcomes in the active arm using only the observed data from the active arm.

To illustrate that RR variance estimator can be positively biased in less extreme cases than that considered above, we carried out a simulation study. We considered a trial with $J = 4$, $n = 200$ and $P(T = r) = P(T = a) = 0.5$. Patients in the active arm deviated (non-informatively) at time 2 ($D = 2$) with probability 0.2; otherwise they did not deviate ($D = 4$). There was no deviation in the reference arm. The treatment effect of interest was $\theta = E(Y_4 \mid T = a) - E(Y_4 \mid T = r)$. For each non-deviating patient in arm T, outcome vector (Y_1, Y_2, Y_3, Y_4) was generated from a normal distribution with mean $\boldsymbol{\mu}_T$ and variance $\boldsymbol{\Sigma}_T$. We used the same mean and variance as in Lu (2014). Specifically, $\boldsymbol{\mu}_r = \boldsymbol{\mu}_a = (29, 22, 17, 14)^T$ for a 'no-treatment effect' scenario, $\boldsymbol{\mu}_a = (29, 20, 14, 11)^T$ and $\boldsymbol{\mu}_r = (29, 22, 17, 14)^T$ for a 'treatment effect' scenario. For both scenarios, the (j, k)th entry of $\boldsymbol{\Sigma}_a = \boldsymbol{\Sigma}_r$ was $36 \times (1 - 0.2 \times |k - j|)$. For deviating patients, (Y_1, Y_2, Y_3, Y_4) was also generated from a normal distribution but with mean and variance depending on the assumed imputation procedure. For example, in the 'treatment effect' scenario, the mean and variance were $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$ for

the 'copy reference' procedure, but $(29, 22, 22, 22)$ and $\boldsymbol{\Sigma}_a$ for the LMCF procedure. Table 1 shows the true values of θ. Note that for the LMCF imputation procedure, $\theta \neq 0$ even when $\boldsymbol{\mu}_a = \boldsymbol{\mu}_r$ (the 'no-treatment effect' scenario).

For each of the two treatment effect scenarios and Carpenter's five imputation procedures, 10000 datasets were generated. The standard analysis-of-covariance (ANCOVA) estimator was first applied to each complete dataset, yielding the complete-data estimator $\hat{\theta}_{\text{comp}}$. Post-deviation outcomes were then discarded and $Q = 1000$ imputed datasets created using the correct imputation procedure (i.e. that assumed when generating the complete data). The ANCOVA estimator was applied to each of these Q imputed datasets, and estimates and standard errors combined using Rubin's Rules, yielding $\hat{\theta}_{(Q)}$ and $\hat{\text{SE}}(\hat{\theta}_{(Q)})$. The *norm* package in R (Schafer, 2012) was used to draw from the posteriors of $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ and $(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$.

Table 1 shows the results. These demonstrate that the RR estimate of the standard error of the treatment effect overestimates the true standard error for the 'copy reference', 'jump to reference' and 'copy increments in reference' procedures. This mirrors findings for the alternative placebo-based pattern mixture model approach presented in Lu (2014). The RR estimator achieves coverage at close to the nominal rate for the LMCF procedure. While conservative variance estimates may sometimes be viewed as desirable, our simulation study highlights another issue with Carpenter et al.'s imputation procedures: they yield smaller empirical standard errors than the estimator based on the complete data. This reflects the strength of the assumption being made by the imputer.

Conclusion

While we welcome Carpenter et al.'s proposals, we are concerned that they may cause bias when deviations are informative (Problem 1). Methods from the causal inference literature (White, 2005) may be helpful to avoid such bias. Problem 2 may be of less practical importance if the reduction in variance caused by making a highly informative assumption like 'jump to reference' is unwanted. If this is so, the positive bias in the RR

variance estimator may balance this reduction, thus yielding a variance estimate that better reflects the real uncertainty. However, it is not clear how this estimate should be interpreted in terms of repeated sampling. Alternatively, one could seek a different variance estimator, e.g. using the general methodology of Robins and Wang (2000). Lu (2013) used the delta method to derive a variance estimator that is consistent under an assumption somewhat similar to 'copy reference'. He also derived a related Bayesian estimator.

References

Carpenter, J.R., Roger, J.H., Kenward, M.G. (2013) Analysis of Longitudinal Trials with Protocol Deviations: a Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. *Journal of Biopharmaceutical Statistics* 23: 1352–1371.

Lu, K. (2014) An Analytic Method for the Placebo-Based Pattern-Mixture Model. *Statistics in Medicine* 33: 1134–1145.

Meng, X.L. (1994) Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 9: 538-558.

Robins, J.M., Wang, N. (2000) Inference for Imputation Estimators. *Biometrika* 87: 113–124.

Schafer, J.L. norm: Analysis of Multivariate Normal Datasets with Missing Values. *R Package Version 1.0-9.4 (Ported to R by Alvaro A. Novo).* http://CRAN.R-project.org/package=norm.

White, I.R. (2005) Uses and Limitations of Randomization-Based Efficacy Estimators. *Statistical Methods in Medical Research* 14:327–347.